AI Accountability Policy Request for Comment by the National Telecommunications and Information Administration

Global Partners Digital submission
12 June 2023

About Global Partners Digital

Global Partners Digital (GPD) works to ensure that human rights underpin the development, use and governance of digital technologies.

GPD has a long history of engagement within digital policy processes at the global, regional and national levels to ensure that frameworks, norms and standards that govern digital technologies are human rights-respecting. This has included efforts to shape AI laws and policies, particularly through our ongoing engagement at the Council of Europe's Committee on AI (CAI), which has been tasked with developing the world's first treaty on AI.

Introduction

We welcome the opportunity to provide comments on the AI Accountability Policy Request for Comment issued by the National Telecommunications and Information Administration.

GPD is pleased that the government is undertaking this effort to interrogate what effective accountability means in AI policy. AI has been compared in recent days with electricity, as something truly revolutionary that will radically change our lives and power each one of our human activities. Whether this claim becomes true is something that remains to be seen, but what is clear is that AI at its current stage of development is not something that happens against our will, and there is a fundamental opportunity for concerted action by governments, companies, and the general public to ensure that AI is regulated and overseen in a way that mitigates and redresses existing harms and is underpinned by transparency.

In this landscape it is useful to remember that AI is yet another form of technology, hence a human creation. This means that existing frameworks and types of regulations which have been developed in previous decades are still relevant today, and we can use these to guide ongoing efforts to harness the advantages of AI technologies and mitigate its harms. This includes critically the human rights international framework.

Questions

AI Accountability Objectives

1. What is the purpose of AI accountability mechanisms such as certifications, audits, and assessments?

Responses could address the following:

a. What kinds of topics should AI accountability mechanisms cover? How should they be scoped?

b. What are assessments or internal audits most useful for? What are external assessments or audits most useful for?

c. An audit or assessment may be used to verify a claim, verify compliance with legal standards, or assure compliance with non-binding trustworthy AI goals. Do these differences impact how audits or assessments are structured, credentialed, or communicated?

d. Should AI audits or assessments be folded into other accountability mechanisms that focus on such goals as human rights, privacy protection, security, and diversity, equity, inclusion, and access? Are there benchmarks for these other accountability mechanisms that should inform AI accountability measures?

e. Can AI accountability practices have meaningful impact in the absence of legal standards and enforceable risk thresholds? What is the role for courts, legislatures, and rulemaking bodies?

Answer Q1: The purpose of AI accountability mechanisms, including both technical and institutional mechanisms, is to ensure that risk assessment - particularly human rights impact assessment- independent audit and traceability of mitigation measures are implemented by entities to evaluate their appropriateness, as an iterative process throughout the entire AI lifecycle.

AI design and deployment should require that any human and environmental harms are appropriately considered to make decisions about the actual implementation of the system, its particular characteristics, and mitigated or remediated in a timely manner. Appropriate mechanisms should be available for grievance (presenting a claim) and effective remedy for individuals and groups adversely affected by AI systems performance.

This should be a focus of regulatory efforts undertaken in different jurisdictions, either by enhancing current mechanisms within existing data protection frameworks, anti-discrimination legislation, consumer protection or other sectoral regulations, or by enacting new rules that assign specific liability to AI systems controllers[1] according to the likelihood and severity of harm that can be caused by their technologies. Providing accountability is part of companies responsibilities under the UN Guiding Principles on Business and Human Rights[2] (UNGPs) anchored in three pillars: protect, respect and remedy.[3]

Accountability mechanisms on AI systems are linked to the internal structures and resources devoted to ensuring a thorough evaluation of the aims pursued with AI system design and deployment. Accountability is built into the AI lifecycle when internal governance structures of the AI systems controllers (either private or public entities) ensure that decision making is made in full awareness of the risk assessment highlighted above, and the process to identify and ponder those risks is documented along AI systems design, development, and deployment. Through these governance processes, entities can ensure better alignment with human rights promotion and protection, resulting in the pursuance of legitimate aims that are appropriate and proportionate to its context of deployment.

Accountability mechanisms should avoid responsibility dilution among the different AI systems controllers that can be different entities during the AI lifecycle. Liability should be clearly and proportionately assigned to the level in which those different entities are best positioned to prevent or mitigate harm in the AI system performance.

Accountability mechanisms available for AI systems in the research and testing phase or business to business might differ from those necessary for massive consumption products. Accountability mechanisms should be able to ensure the necessary safety and quality assurance processes that a massive consumption product needs to avoid consumer harm once in the market or provide mechanisms of remedy for impacted consumers when those measures have not been taken.

Impact or risk assessment are the frameworks more often referred to address the evaluation of likelihood and severity of the deployment and use of AI systems to produce negative consequences at the individual or collective level in the social context of its inception.

[1] This submission uses “controllers” and “entities” interchangeably to refer to those are in the position to influence the design, development or deployment of AI systems during its lifecycle.

[2] See https://www.ohchr.org/Documents/Publications/GuidingPrinciplesBusinessHR_EN.pdf

[3] Human rights due diligence is grounded in the UN Guiding Principles on Business and Human Rights endorsed by the UN Human Rights Council in 2011. The UNGPs prescribe that businesses should respect human rights by performing human rights due diligence (Guiding Principle 17), establishing a process for identifying and assessing adverse human rights impacts with which a business is involved (Guiding Principle 18), preventing and mitigating these impacts (Guiding Principle 19), tracking them (Guiding Principle 20), and communicating them (Guiding Principle 21).

Impact assessment provides opportunity for a reflexive process in which internal -and ideally external stakeholders beyond those in charge of design, development, and deployment of AI systems- can engage with a wide range of socio-technical aspects of technology and document their observations, predictions and measures taken to mitigate the identified risks.

Human rights due diligence refers to the assessment of the likelihood and severity of the AI system to cause or contribute to potential and actual impacts on the exercise of civil, political, economic, social, cultural and environmental rights -recognised as part of the international human rights framework- through their own activities, or directly linked to their operations, products or services or their business relationships. It provides a concrete frame of reference to identify the possible impacts and measure them, relaying in the experience of 75 years of interpretation and application of the international human rights framework that are part of the international legal obligations of states parties.[4]

From a methodological perspective, to use a due diligence perspective means that the impact assessment is considered an ongoing process that should take place in every stage of the AI lifecycle, and it is intended to provide opportunity to use its results to continuously assess the legitimacy, necessity and proportionality of the technical deployment, as well as introduce the adequate mitigation measures to ensure the alignment of the use of the technology with human rights realisation.

For ensuring legitimacy and effectiveness of the exercise it is crucial ensuring the proper documentation of findings and the communication of its results allowing the external scrutiny of the process. Human rights due diligence should benefit from a wide range of technical and social science expertise contributions, including human rights experts, and other sectoral experts, community members, and representatives of marginalised and vulnerable groups who hold a unique perspective by living with the harms created or exacerbated by AI systems.

Human rights impact assessments (HRIAs) are born from UNGPs Principle 18 referenced above, and can be described as "a process for identifying, understanding, assessing and addressing the adverse effects of a [project, product, services, or activities] on the human rights enjoyment of impacted rightsholders."[5] Some HRIAs are ex ante assessments of a business or technology's potential impacts, and others provide an authoritative ex post assessment of an AI system's actual impacts. Currently, there is no single prescribed process or method for conducting HRIAs for

[4] See OHCHR. UN Human Rights "Issues Paper" on legislative proposals for mandatory human rights due diligence by companies. June 2020. Available at: https://www.ohchr.org/sites/default/files/Documents/Issues/Business/MandatoryHR_Due_Diligence_Issues_Paper.pdf

[5] Danish Institute for Human Rights. 2020. Guidance on Human Rights Impact Assessment of Digital Activities. Available at: https://www.humanrights.dk/publications/human-rights-impact-assessment-digital-activities

AI systems. HRIA methodologies must be adapted to best fit the needs of external stakeholders and must be responsive to the specific contexts.[6]

The responsibility to perform human rights due diligence or HRIAs to assess risks of AI systems design, development and deployment should rest in those in better position for identifying the risks and harms during each stage of the AI lifecycle. This means that not only designers but also deployers of AI should be responsible for conducting this ongoing evaluation process and to communicate to each other, to the impacted communities, the oversight authorities, and the general public the results of the assessment exercise.

According to UNGPs, human rights due diligence or HRIAs critically require ensuring meaningful participation in the risk identification and comments about the impacts, its severity and likelihood, and development of harm prevention and mitigation measures from potentially affected groups and other relevant stakeholders in the context of implementation of the AI system under evaluation.

2. Is the value of certifications, audits, and assessments mostly to promote trust for external stakeholders or is it to change internal processes? How might the answer influence policy design?

Answer Q2: Any assessment process within an accountability system serves a dual purpose of helping technology systems controllers to identify risks and act to prevent them by improving their internal processes and the technology structure itself, but also if they are adequately documented and externally shared with external stakeholders this would support the creation of trust which is not an intrinsic value but rather the result of having a proper systems of accountability in place. This radically influences policy design because beyond the responsibilities assigned to build an accountability system and to perform assessment as part of it, there should be considered the transparency layer about how, by whom, under which standards and with which results are the assessments conducted.

3. AI accountability measures have been proposed in connection with many different goals, including those listed below. To what extent are there tradeoffs among these goals? To what extent can these inquiries be conducted by a single team or instrument?
a. The AI system does not substantially contribute to harmful discrimination against people.

b. The AI system does not substantially contribute to harmful misinformation, disinformation, and other forms of distortion and content related harms.

c. The AI system protects privacy.

[6] See ECNL and Data & Society. Recommendations for Assessing AI Impacts to Human Rights, Democracy, and the Rule of Law, November 2021. Available at: https://ecnl.org/sites/default/files/2021-11/HUDERIA%20paper%20ECNL%20and%20DataSociety.pdf

d. The AI system is legal, safe, and effective.

e. There has been adequate transparency and explanation to affected people about the uses, capabilities, and limitations of the AI system.

f. There are adequate human alternatives, consideration, and fallbacks in place throughout the AI system lifecycle.

g. There has been adequate consultation with, and there are adequate means of contestation and redress for, individuals affected by AI system outputs.

h. There is adequate management within the entity deploying the AI system such that there are clear lines of responsibility and appropriate skillsets.

Answer Q3: The goals listed in the question can be subsumed without contradiction in the human rights approach that we highlighted in our answer to Q1. When there is potential for clashing in the fulfilling of different fundamental rights in place, there is guidance coming from the international human rights framework to balance them in a way in which it is possible to ensure their exercise in a proportionate manner and with the least limitations to make them compatible.

As highlighted in Q1, the participatory nature of the AI decision making and assessment process is key in order to ensure the effectiveness and the legitimacy of the technology.

Accountability can only truly exist when there are available mechanisms for grievance and remedy that have been integrated as part of the governance structure of the technology at stake. Building proper governance processes inside private or public institutions in charge of designing, deploying or developing AI systems requires leadership commitment of the adequate resources, skills sets and time to perform the assessment process and meaningful responses to integrate the result of those processes into the technology lifecycle.

4. Can AI accountability mechanisms effectively deal with systemic and/or collective risks of harm, for example, with respect to worker and workplace health and safety, the health and safety of marginalized communities, the democratic process, human autonomy, or emergent risks?

Answer Q4: According to what has been highlighted in our answer to Q1, we believe that accountability mechanisms built around a human rights approach are able to deal with systemic, collective and individual risks and harms.

For that purpose, technology needs to be designed and deployed in an economic and socially inclusive manner to counter the harms of a disruptive technology. Addressing bias in data sets used for training the AI system is just one part of the inclusiveness issue related with the representativeness of the data and the quality of outputs of the system when deployed in a specific context. But another essential part of inclusiveness is dealing with the consequences for the larger human ecosystem that the AI system will impact, such as availability of employment, ability to access to social welfare, healthcare or other relevant services.

Open, transparent, and participatory design, development and deployment of AI systems require that a broad range of perspectives and interests should be taken into account, reflecting differences in culture, language, expertise, socio-economic conditions, and vulnerability conditions to ensure its inclusiveness.

In many cases, the focus is inadequately put on equal access to AI technology, the ability for everyone to benefit from the AI deployment, but this can only be truly possible if the decision making over design and deployment of a certain AI technology and its iterative evaluation considers the participation of a wide range of stakeholders, including those that are impacted by the system deployment and not only those controlling the system. Otherwise, AI design and deployment may further entrench marginalisation and vulnerability already present in society. Any AI system deployment should take into consideration current inequalities in access to infrastructure and knowledge, and provide meaningful mitigation for impacted populations to ensure alignment of AI with a full exercise of their human rights. Beyond that, assessment of AI deployment should consider accessible alternatives for those unable or unwilling to access technology to avoid further marginalisation.

Resources should be devoted by designers and deployers of AI systems to monitor and mitigate disproportionate impacts on specific groups of people that reflect bias and discrimination in the outputs of the system. Mechanisms for redress should be made available in those cases. For that purpose, existing legal frameworks dealing with non-discrimination in different fields (such employment access, consumer or healthcare) should be strengthened and used to guide AI systems deployed, or created in jurisdictions or sectors in which they do not exist.

5. Given the likely integration of generative AI tools such as large language models (e.g., ChatGPT) or other general-purpose AI or foundational models into downstream products, how can AI accountability mechanisms inform people about how such tools are operating and/or whether the tools comply with standards for trustworthy AI?

Answer Q5: The recent release of generative AI systems for mass consumption by the public has prompted significant concern and attention on AI capabilities and potential for risk. The availability of these tools for public consumption has spurred discussion on the risk of automation bias, which refers to the tendency of humans to consider automated systems more capable, objective, and reliable than humans in performing a task. For the human understanding, it is counterintuitive that the synthetic language provided by these systems is deprived of contextual meaning and communicational intent, as it is implicit in language used by humans. They are simple fabrications based on the likelihood of language strings sequences learnt from its training data.

Part of this trend is fuelled by the fact that the release of these tools has not been accompanied with sufficient efforts of consumer transparency regarding the actual capabilities and limitations of the models. Something that has been painfully experienced by users with deceiving quality of outputs that include nonexistent paper citations or court precedents, and fabrication of profiles or facts.

Given the characteristics described above related with consumer facing generative AI systems, accountability mechanisms should be able to ensure the necessary safety and quality assurance processes that a massive consumption product needs to avoid consumer harm once in the market or provide mechanisms of remedy for impacted consumers when those measures have not been taken.

Foundational models providers developed upstream in the value chain should be responsible for collaborating with implementers in identifying additional potential harms that specific context implementation could have. The liability regime established by the accountability regime should account for the way in which developers of foundational models and implementers should contribute to remedy in case of harm.

6. The application of accountability measures (whether voluntary or regulatory) is more straightforward for some trustworthy AI goals than for others. With respect to which trustworthy AI goals are there existing requirements or standards? Are there any trustworthy AI goals that are not amenable to requirements or standards? How should accountability policies, whether governmental or non-governmental, treat these differences?

Answer Q6: The design, development and deployment of AI is underpinned by a range of expectations about its ability to meaningfully contribute to overcome inequality, access to services and economic development under the premise that what is needed is to allow this innovation to grow and provide policies that are flexible enough to unleash its advantages and help us to achieve our sustainable development goals. Trustworthiness is usually identified as an AI intrinsic value that should be fostered through design. But trustworthiness is the ability to be relied on as fair, just or truthful. It is rather a result than an intrinsic value, and it is a product of reaching benchmarks and fulfilling expectations.

In that sense, accountability measures in the form of AI technology governance (whether voluntary or mandatory) have to point out the elements that are necessary to ensure fairness and human rights realisation through the technology deployment. Relevant standards have been built over the recent years around cybersecurity, privacy and non-discrimination (in different fields such as employment, financial services, insurance, health) that are still applicable and useful for guiding AI implementation.

What is currently missing and deserves public investment in developing research is the building of risk measurement standards that are specifically applicable to AI systems implementation in different fields. A relevant challenge that any accountability framework confronts is to be able to clearly set expectations about the type of benchmark that should be fulfilled by the AI systems controllers in order to consider that they are achieving the threshold for creating trustworthiness.

7. Are there ways in which accountability mechanisms are unlikely to further, and might even frustrate, the development of trustworthy AI? Are there accountability mechanisms that unduly impact AI innovation and the competitiveness of U.S. developers?

Answer Q7: Accountability mechanisms are the cornerstone of responsible technology development. The creation of such mechanisms, including through regulation and voluntary measures, will bolster the leading role of US developers by building in strong incentives and the ability to compete across jurisdictions on a global level. This is particularly relevant as other governments and regional blocs are actively pursuing their own forms of AI governance. It is expected that US developers that are able to integrate those accountability mechanisms to their technology R&D will benefit from the ability to better ensure regulatory compliance in places in which those regulations become mandatory, as is the case of relevant markets such as Europe or Brazil that are currently discussing mandatory regulations.

Accountability Subjects

15. The AI value or supply chain is complex, often involving open source and proprietary products and downstream applications that are quite different from what AI system developers may initially have contemplated. Moreover, training data for AI systems may be acquired from multiple sources, including from the customer using the technology. Problems in AI systems may arise downstream at the deployment or customization stage or upstream during model development and data training.

a. Where in the value chain should accountability efforts focus?

b. How can accountability efforts at different points in the value chain best be coordinated and communicated?

c. How should vendors work with customers to perform AI audits and/or assessments? What is the role of audits or assessments in the commercial and/or *public procurement process? Are there specific practices that would facilitate credible audits (e.g., liability waivers)?*

d. Since the effects and performance of an AI system will depend on the context in which it is deployed, how can accountability measures accommodate unknowns about ultimate downstream implementation?

Answer Q15: Accountability needs to be embedded throughout the whole value chain, or more specifically, throughout the entire lifecycle of the AI system. In that sense, the responsibility for risk management and harm reduction should lie with the AI systems controllers during its design, deployment, and development as they are best positioned to address them. As stated above, the risk assessment is an iterative exercise that should be performed throughout the AI lifecycle.

There is an inherent responsibility of all entities participating in the AI system lifecycle to document and communicate to each other the activities conducted to assess and mitigate risks. The type of activities conducted in each stage could vary in nature and intensity, but should be more exigent when the AI system is intended to be made available for public utilisation.. There are fundamental differences on the risk tolerance that can be permissible when the AI systems still are in the research and development stage, from those necessary to implement when the public interacts with an AI system. This by no means implies that vendors can disengage from the harmful results that an AI system can produce when it is at later stages of the AI lifecycle, including its deployment and other downstream uses. The assessments and audits performed later on in the AI system lifecycle should be intended to give feedback and serve to provide corrections and improvements that are implemented by the vendor. The responsibility of the vendor and accordingly transparency that needs to be offered is linked to the ability of the vendor to control the training dataset responsible in great measure of the AI system performance. For all these reasons, liability waivers do not seem appropriate, and there is a clear need for a dynamic distribution of the legal liability in case of harm.

As highlighted in the question, the implementation stage is an essential context in which to assess the risk of harm and to identify adequate preventive and mitigation measures. There is accordingly a high level of responsibility of the AI system designers and deployers acting as vendors of these technologies to disclose he training dataset used by them or at least their characteristics, in order to allow the downstream implementers to better assess the appropriateness of the AI systems for their context implementation and the need for additional training and risk assessment before deploying the system, or even be able to decide abandoning the implementation.

In the field of public procurement, it is imperative that the state establish high requirements of transparency of vendors regarding a particular AI system, its data training conditions, the quality assurance measures adopted, the risk assessment performed, the mitigation measures implemented and the external audits performed.

16. The lifecycle of any given AI system or component also presents distinct junctures for assessment, audit, and other measures. For example, in the case of bias, it has been shown that "[b]ias is prevalent in the assumptions about which data should be used, what AI models should be developed, where the AI system should be placed—or if AI is required at all." How should AI accountability mechanisms consider the AI lifecycle? Responses could address the following:

a. Should AI accountability mechanisms focus narrowly on the technical characteristics of a defined model and relevant data? Or should they feature other aspects of the socio- technical system, including the system in which the AI is embedded? When is the narrower scope better and when is the broader better? How can the scope and limitations of the accountability mechanism be effectively communicated to outside stakeholders?

b. How should AI audits or assessments be timed? At what stage of design, development, and deployment should they take place to provide meaningful accountability?

c. How often should audits or assessments be conducted, and what are the factors that should inform this decision? How can entities operationalize the notion of continuous auditing and communicate the results?

d. What specific language should be incorporated into governmental or non-governmental policies to secure the appropriate timing of audits or assessments?

Answer Q16: Accountability mechanisms should be a combination of socio-technical measures, oriented to create a governance regime that can account for the context of AI system deployment. The technical characteristics and the identification of the data necessary to build the system cannot be separated from its deployment or implementation. Many of the harms which have resulted from the deployment of AI systems in recent years were due to the fact that the AI system was originally intended for a particular context, but was then deployed in another without careful consideration of the differences in terms of institutional frameworks, impacted populations and representativeness of the data set.[7]

Audit or assessment should be undertaken throughout the AI lifecycle with different purposes to serve as an effective accountability mechanism. During the design and development of the AI system, the audit or assessment should aim to identify the risks of the technology to cause negative impacts to the exercise of fundamental rights and freedoms. This should result in the ability to insert into the dataset curation or provide the necessary measures to prevent or mitigate those risks, or to decide against deployment where it is not possible to mitigate such risks. During the

[7] Eubanks, Virginia, Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor, St. Martin Press, New York, 2018. Derechos Digitales, Inteligencia Artificial e Inclusion. Study Cases. Available at: https://ia.derechosdigitales.org/en/

deployment and use of the AI system, the audit or assessments should verify how the system operates in the relevant context and thus inform the need for corrective measures to ensure ongoing prevention or mitigation of risks. This assessment may ultimately result in a decision to halt the deployment when the harms outweigh the benefits of the system. For this purpose, language that accompanies the requirement for audits or assessments should reference the idea of an ongoing iterative process, intended to give feedback in the design and functioning of any AI system.

In order for audit and assessment to be an effective part of the accountability regime, there is need for relevant entities to adequately document these processes, practised by independent actors, and to the biggest extent possible make them available for external access or accompanied by transparency reports widely available to public and impacted groups.

17. How should AI accountability measures be scoped (whether voluntary or mandatory) depending on the risk of the technology and/or of the deployment context? If so, how should risk be calculated and by whom?

Answer Q17: There is a fundamental need to advance in the development of benchmarks for risk measurement considering both elements: the specific field of AI deployment and the implementation context. There are some elements of accountability measures that are common to the risk assessment process no matter what the specific field of implementation is. For example, risk assessment should consider a wide range of stakeholder contributions to the identification of risks, including multidisciplinary experts, concerned community and vulnerable groups likely to be impacted by the AI system.

When considering who should be responsible for measuring risks, the principle should be that the best positioned entity to define and control the AI system should be the one responsible for the measurement considering the contributions of the external stakeholders referenced above. The development of benchmarks for risk assessment should provide an opportunity to assess the quality and pertinence of the risk assessment performed by the AI system controllers, and transparency about the assessment process and its result should facilitate the external evaluation of it.

18. Should AI systems be released with quality assurance certifications, especially if they are higher risk?

Answer Q18: Quality assurance certifications could be one way to ensure the performance and quality of risk assessments before deployment of an AI system. There are some fields where AI systems may be used that already have a long tradition of certification, in which they will be consistent with the requirements of other technical systems that are currently subject to strict control. This is, for example, the case of AI systems implemented in healthcare that could warrant a level of certification given the sensitivity of dealing with human health. Other fields of deployment such as employment, predictive policing, military use or migration control are additional examples of areas with higher risk that could deserve some kind of certification system as part of accountability mechanisms.

19. As governments at all levels increase their use of AI systems, what should the public expect in terms of audits and assessments of AI systems deployed as part of public programs? Should the accountability practices for AI systems deployed in the public sector differ from those used for private sector AI? How can government procurement practices help create a productive AI accountability ecosystem?

Answer Q19: The deployment of AI systems by governments should be subject to the highest standards in terms of risk assessment and accountability mechanisms in place, in order to ensure democratic participation in the decision making of technology implementation, as well in all the process from procurement to deployment of the AI systems. There is a fundamental responsibility from the states to ensure that any technology deployment is consistent with the promotion and protection of human rights. There is also a relevant role to play from the government in setting the example in terms of responsible deployment of AI systems that can have a positive impact in the deployment of responsible practices from the AI industry.

Accountability Inputs and Transparency

21. What are the obstacles to the flow of information necessary for AI accountability either within an organization or to outside examiners? What policies might ease researcher and other third-party access to inputs necessary to conduct AI audits or assessments?

Answer Q21: The obstacles associated with the flow of information in accountability mechanisms are related with the information asymmetry between different internal actors of the organisation designing and implementing the AI systems, and the asymmetries of information confronted by external actors engaged in the assessment and oversight of the AI systems.

Overcoming the internal asymmetries of information require relevant governance mechanisms to ensure that there is involvement in the design and decision making about the systems from different types of expertise inside the AI system vendor. From a human rights approach, this implies that AI system design and development is subject to socio-technical assessment, and that leadership of the organisation has involvement in the decision making and meaningful information to address the human rights impacts coming from the technology.[8]

[8] A good example of good governance practices, although linked with a limited set of human rights, are part of the Global Network Initiative Principles and implementation Guidelines, available here: https://globalnetworkinitiative.org/gni-principles/

From the perspective of the external stakeholder asymmetries, good practices or regulations that can ensure the access for vetted researchers to commercially sensitive information could be a way in which the asymmetry could be confronted without harming the business interests. In this matter it is worth attention the mechanism established recently by the article 40 of the Digital Service Act in the European Union, and the way in which the mechanism will unfold in its implementation.

22. How should the accountability process address data quality and data voids of different kinds? For example, in the context of automated employment decision tools, there may be no historical data available for assessing the performance of a newly deployed, custom-built tool. For a tool deployed by other firms, there may be data a vendor has access to, but the audited firm itself lacks. In some cases, the vendor itself may have intentionally limited its own data collection and access for privacy and security purposes. How should AI accountability requirements or practices deal with these data issues? What should be the roles of government, civil society, and academia in providing useful data sets (synthetic or otherwise) to fill gaps and create equitable access to data?

Answer Q22: The issue of data quality should be central in the design of the risk assessment process as part of the accountability regime. As stated above, the identification of the dataset appropriate for the AI systems design and deployment should have fundamentally into consideration the context of implementation in order to avoid the harm coming for the bias and lack of representativity of data for the specific deployment.

The role of government should be to support the research necessary to create good benchmarks about data quality in different fields, and the creation of data commons that fulfil data protection and privacy requirements that can be publicly accessible for AI developers and deployers to ensure AI design and implementation that are aligned with the exercise of fundamental rights. Finally, the government has a fundamental role in ensuring that its own procurement practices are consistent with high standards of dataset quality assurance for the specific context of AI systems implementation inside public administration.

23. How should AI accountability "products" (e.g., audit results) be communicated to different stakeholders? Should there be standardized reporting within a sector and/or across sectors? How should the translational work of communicating AI accountability results to affected people and communities be done and supported?

Answer Q23: Effective communication of risk assessment to stakeholders is key for ensuring the legitimacy of the AI implementation, and the ability of external actors to challenge the results of the assessment process or using them to exercise their right to remedy in the case of harm from the AI deployment. The meaningful transparency of the accountability system in place might imply different levels of access to information for different sectors according to their particularities, but also can require different levels of access to information for different stakeholders.

Without prejudice of sectoral reporting practices that could be developed, it would be useful to have an international effort to standardise the communication of risk of AI systems (e.g., akin to nutritional labelling efforts) that could more easily communicate risks, AI limitations and adopted guardrails for the case of consuming facing products.

Barriers to Effective Accountability

25. Is the lack of a general federal data protection or privacy law a barrier to effective AI accountability?

Answer Q25: Yes. The lack of a comprehensive data protection framework makes it more difficult to set effective accountability mechanisms that can include proper safeguards for privacy, ensure quality assurance of datasets used for the training of AI systems, and that the information collected by AI systems during its deployment can ensure adequately the individuals control over their own personal data. Even though there are sectoral regulations in place that can be leveraged and are fully applicable to AI systems deployment, the lack of uniformity in the rules for different fields negates the possibility of creating risk assessment benchmarks and examples of good practices that could be applicable across sectors.

Moreover, the absence of a comprehensive data protection framework has resulted in a lack of readily available remedies in cases where the deployment of AI systems has a negative impact on individuals' privacy. Even though the FTC has authority to prosecute cases to protect consumers interest from companies harms, the likelihood of being able to gather the data to prosecute those cases is limited for the asymmetry of information around AI systems functioning. Having clear cross-cutting obligations about transparency in the use of AI systems intervention in decision-making and provenance of datasets used in the systems are issues that are currently addressed by the most advanced data protections frameworks in place somewhere else in the world (such as the European GDPR) and the US could benefit from adopting similar rules.

26. Is the lack of a federal law focused on AI systems a barrier to effective AI accountability?

Answer Q26: A federal law could be an effective way to ensure the baseline for accountability regime to ensure the design, development and deployment of responsible AI. However, the absence of a federal law should not be necessarily seen as a complete barrier because there are sectoral regulations in place that are fully applicable to AI systems (such as employment, financial services, insurance, healthcare) that could be applied to enforce the responsibility of AI implementers.

A federal law focused in AI systems could advance the more granular elements of an accountability system in order to better inform the internal process of AI design and development, but this also can be advanced in the meantime with codes of practice and voluntary measures developed by industry or guided by the authority such NTIA has done with its AI Risk Management Framework.

28. What do AI audits and assessments cost? Which entities should be expected to bear these costs? What are the possible consequences of AI accountability requirements that might impose significant costs on regulated entities? Are there ways to reduce these costs? What are the best ways to consider costs in relation to benefits?

Answer Q28: The cost of audits and assessments should be considered the regulatory cost that needs to be internalised by the designers, developers and implementers of AI. This is no different from any other industry that has to bear the cost of dealing with hazardous products. The clearer the standards for the assessments or audits performance can be that will imply benefits in terms of compliance cost for AI systems.

Responsible AI design, development and implementation can be considered a long term investment because it will enable broader regulatory compliance for US AI vendors that would like to offer their products and services to other markets with increasing likelihood of regulatory requirements in place.

29. How does the dearth of measurable standards or benchmarks impact the uptake of audits and assessments?

Answer Q29: As highlighted above, this is a relevant issue which demands urgent public investment in research. The lack of measurable standards or benchmarks creates the risk of rendering impact assessments as unproductive exercises by providing an appearance of accountability but not enough to achieve it effectively. In that sense, the legitimacy of the accountability system and the trustworthiness of AI systems themself rely on the fact that it is possible to assess the quality of its evaluation process by independent actors and the society at large.

As also highlighted above, standards and benchmarks are useful for facilitating compliance and reducing its cost and with that making it easier for new entrepreneurs to compete in the market.

AI Accountability Policies

30. What role should government policy have, if any, in the AI accountability ecosystem? For example:

a. Should AI accountability policies and/or regulation be sectoral or horizontal, or some combination of the two?

b. Should AI accountability regulation, if any, focus on inputs to audits or assessments (e.g., documentation, data management, testing and validation), on increasing access to AI systems for auditors and researchers, on mandating accountability measures, and/or on some other aspect of the accountability ecosystem?

c. If a federal law focused on AI systems is desirable, what provisions would be particularly important to include? Which agency or agencies should be responsible for enforcing such a law, and what resources would they need to be successful?

d. What accountability practices should government (at any level) itself mandate for the AI systems the government uses?

Answer Q30: A baseline federal regulation could be beneficial to set the main elements for accountability mechanisms that should be put in place. This could be combined with more granular guidance with a sectoral approach that can provide attention to the particularities of specific fields of application.

As stated above, a central element of any accountability regime should be addressing the information asymmetries in order to enhance the external stakeholder assessment and the authority oversight of the quality of the evaluation performed of the AI system. This includes the documentation trail of the assessment, reporting on the way the assessment has been performed, and some degree of transparency of the above to be accessed by external actors (maybe with differential levels of accessibility).

As important as the focus on transparency around the elements of the AI system and its assessments, it is the element of grievance and remedy inside the accountability system. This should be covered by AI regulation by mandating baseline elements of accessible claim systems available for AI systems impacted individuals and groups. Regarding remedy, the regulation should ensure to assign liability of AI systems controllers in each stage of the AI lifecycle in the proportion they have or have not effectively contributed to the effective prevention or mitigation of harm. In that sense, the system should facilitate the proof of claims for impacted individuals or groups by assigning joint liability to the controllers but allow them to demonstrate how effective they have been in assessing and mitigating risk in order to reduce their responsibility.

Finally, to ensure compliance with the AI regulation, authority should be given to an agency or body sufficiently independent and resourced to ensure effective oversight.

31. What specific activities should government fund to advance a strong AI accountability ecosystem?

Answer Q31: As stated above, there is a need for relevant investment from the government to support research for creation of risk assessment standards or benchmarks in order to ensure the effectiveness of AI risk assessment.

Another area that would benefit from strong government investment is the production of guidelines and best practices for government and companies around meaningful multi-stakeholder participation in the AI assessment process. This will allow us to break the current silos in AI development to provide a more comprehensive socio-technical discussion, which is a more consistent approach to infuse democratic values to the development and deployment of AI.

Finally, public investment should be given to the creation of data infrastructure to ensure the representativeness and quality of data sets in different settings of AI design and deployment with a view to avoid bias and discrimination. This includes also the development of public infrastructure and computing power allowing access for entrepreneurs and social innovators to avoid the market closure in the hands of a few current dominant actors that often lack diversity and alignment with democratic values.

34. Is it important that there be uniformity of AI accountability requirements and/or practices across the United States? Across global jurisdictions? If so, is it important only within a sector or across sectors? What is the best way to achieve it? Alternatively, is harmonization or interoperability sufficient and what is the best way to achieve that?

Answer Q34: As technology development does not recognise physical frontiers, there is a fundamental value to ensure that AI system design, development and deployment is subject to consistent standards across jurisdictions. That is why in our answers we propose to use the international human rights framework as the guidance for AI risk assessment and regulation.

Maria Paz Canales
Head of Legal, Policy and Research
12 June, 2023

Ian Barber
Legal Lead
12 June, 2023

Ian Barber